Collaborative Xmeans-EM Clustering for automatic detection and segmentation of moving objects in video

Hiba Ramadan, Hamid Tairi
LIIAN Laboratory, Department of Informatics Faculty of Sciences Dhar-Mahraz
University of Sidi Mohamed Ben Abdellah
Fez, Morocco
ramadanhiba@gmail.com, htairi@yahoo.fr

Abstract—Detecting and segmenting moving objects in video is a challenging and essential task in a number of applications. This paper presents a new algorithm of moving objects detection and segmentation. Firstly, we extract Selective Spatio-Temporal Interest Points (SSTIPs). The next step is to partition the SSTIPs into a set of moving clusters. To reduce the impact of the choice of a clustering method and its parameters on the quality of the result, we propose to integrate the concept of collaborative clustering of two clustering algorithms without requiring a user-defined number of clusters: Xmeans and Expectation–Maximization (EM) clustering. Finally, the segmentation of the objects associated to the given clusters is performed using an automatic maximal similarity based region merging (MSRM) method. Our algorithm is evaluated on several sequences and experimental results show a good performance for automatic detection and segmentation of moving objects.

Keywords—moving objects detection; Selective Spatio-Temporal Interest Points; collaborative clustering; segmentation; automatic maximal similarity based region merging

I. INTRODUCTION

Moving objects detection in video sequences is a challenging problem. It has application in numerous fields such as video-surveillance, robotics etc.

Various methods exist in the literature to detect and segment moving objects. Good reviews of the different techniques can be found in [1], [2] and [3].

In this paper, we propose a new algorithm to detect and segment moving objects based on the clustering of the Selective Spatio-Temporal Interest Points (SSTIPs) [4] by applying the collaborative clustering [5] of two clustering algorithms and segmenting automatically the detected clusters using the maximal similarity based region merging (MSRM) method [6] .

II. MATERIALS AND METHODS

A. Selective Spatio-Temporal Interest Points (SSTIPs)

Existing STIPs detectors are vulnerable to camera motion and moving background in videos, therefore they detect unwanted STIPs in the background. To overcome these challenges, the author of [4] introduces a novel approach for selective STIPs (SSTIPs) detection, by first detecting the spatial interest points (SIPs), then applying surround suppression combined with local and temporal constraints, achieving robustness to camera motion and background clutter.

The strong aspect of the SSTIPs detector is, it can detect dense STIPs at the motion region without affected by the complex background. This is an important property to perform the detection of the moving objects by clustering and the initialization of the segmentation.

A. Collaborative Clustering

Different methods can, from the same data, produce very different clustering results. Furthermore, even the same algorithm can produce different results, according to its parameters and initialization. However, different studies showed that the combination and the collaboration of different clusterings may increase their efficiency and accuracy

SAMARAH method proposed in [5] uses an unsupervised collaborative multi-strategy process to enhance classification. Its principle consists of an automatic and mutual refinement of the initial clustering results, until all the results have almost the same number of clusters, and all the clusters are statistically similar with a good internal quality. At the end of this process, as the results have comparable structures, it is possible to define a correspondence function between the clusters, and to apply a unifying technique, such as a voting method.

B. MSMR Segmentation

MSMR [6] is an interactive region merging method where the interactive information is introduced as markers to roughly indicate the position and main features of the object and background. Then the method will calculate the similarity of different regions and merge them based on the maximal similarity rule, which is adaptive to image content and does not require a preset threshold.

Firstly, a pre-segmentation step is necessary using a low level segmentation method, then the region merging process is conducted by merging a region into one of its neighboring regions, which has the most similarity. The color histogram was chosen as an effective region descriptor and the similarity measure is given by:

978-1-5090-0478-2/15/$31.00 ©2015 IEEE

$$\rho(R,Q) = \sum_{\mu=1}^{4096} \sqrt{Hist_R^{\mu}.Hist_Q^{\mu}} \qquad (1)$$

Where $Hist_R$ and $Hist_Q$ are the normalized histogram of the regions R and Q.

III. PROPOSED METHOD

Our algorithm of automatic detection and segmentation of moving objects begins with the extraction of SSTIPs. Next, we need to define the features witch will be used for create the moving clusters.

Similarly to [3], we will use three groups of features: the spatial position, the motion and the photometry.

To characterize the displacement of the SSTIPs, we have to compute the optical flow. We choose to apply an accurate and new model based on the correlation transform [7], this method gives a dense and a robust optical flow, even in sequences where severe illumination changes are present.

To characterize the photometry, we need to extract the color information of each SSTIP. Here, we use the HSV color space instead of the RGB color one witch does not separate luminance and chrominance, and the R, G, and B components are highly correlated. So, we describe each SSTIP $s(x, y)$ with its hue value computed over a neighbourhood on a 3x3 window to be robust to noise. Also, to include some simple temporal consistency, we add the hue value at time $t + 1$ for the displaced point $s'(x + dx, y + dy)$.

Once we have defined each SSTIP descriptor, we address the problem of grouping the SSTIPs into clusters.

Our first contribution here is to apply the collaborative clustering, without requiring a user-defined number of clusters, to the Xmeans algorithm [8] and the Expectation-Maximization clustering [9] using cross-validation, in order to enhance the partitioning of the SSTIPs.

To get the complete masks of objects, a final step is necessary: segmenting the object associated to a given cluster.

The MSRM method is simple yet powerful as interactive schema for static image segmentation. Our second contribution in this work, is to apply the MSRM automatically to segment moving objects.

Based on the moving clusters detected using the collaborative Xmeans-EM clustering of SSTIPs, we will initialise the background and the objects markers. So, the SSTIPs associated at each cluster will be considerate as the object seeds and to mark the background we will surround the given cluster with a rectangle.

Each object associated at a given cluster will be segmented separately and the segmentation of all the objects can be performed simultaneously and in parallel.

IV. EXPERIMENTAL RESULTS

One of many applications of our algorithm is human detection and segmentation in video.

Fig.1. presents an example of the results obtained by applying our algorithm in the pedestrian sequence used in [10].

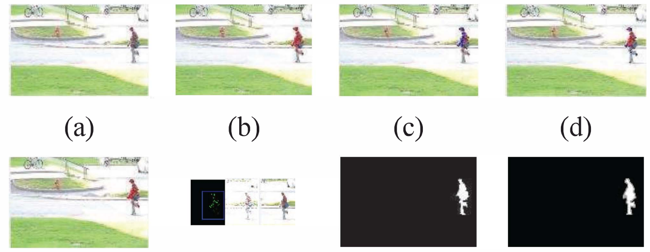

Fig.1. An example from the pedestrian sequence of the automatic detection and segmentation of the moving objects.(a) frame 864 (b) the extracted SSTIPS (c) Xmeans clustering (BIC-Value : -36.353836; two detected clusters) (d) EM clustering (two clusters selected by cross validation) (e) Collaborative Xmeans-EM clustering (one detected cluster) (f) Automatic MSMR segmentation (background and foreground seeds initialization; initial pre-segmentation; MSMR segmentation) (g) the final mask of the moving object (h) the groundthrut

V. CONCLUSION

A new algorithm to detect and segment automatically moving objects in videos has been presented in this paper.

Our approach is based on the Collaborative Xmeans-EM clustering of the SSTIPs for, first, detecting the location of the moving objects without a user interaction, and then, initializing the seeds to perform the automatic MSMR segmentation to obtain the final masks of the objects.

REFERENCES

[1] A. Yilmaz, O. Javed, and M. Shah, “Object tracking: A survey,” ACM Computing Surveys, vol. 38, Dec. 2006 .

[2] K . Bahadir, “Review And Evaluation Of Wellnown Methods For Moving Object Detection And Tracking In Videos,” Journal of Aeronautics and Space Technologies, July 2010.

[3] A. Bugeau, and P. Pérez. “Detection and segmentation of moving objects in highly dynamic scenes,”, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’07), 2007.

[4] B. Chakraborty, Michael B. Holte, Thomas B. Moeslund and Jordi Gonzàlez, "Selective Spatio-Temporal Interest Points," Computer Vision and Image Understanding, vol. 116, pp. 396-410, 2012.

[5] C. Wemmert and P. Gancarski, “A multi-view voting method to combine unsupervised classifications,” Artificial Intelligence and Applications, pp. 447–452, 2002.

[6] J. Ning, L. Zhang , D. Zhang and C. Wu, “Interactive image segmentation by maximal similarity based region merging,” Pattern Recognition, vol 43, pp.445-456, 2009.

[7] M. Drulea and S. Nedevschi, "Motion Estimation Using the Correlation Transform," Image Processing, IEEE Transactions on, vol.22, pp.3260-3270, 2013.

[8] A. W. Moore and D. Pelleg, “X-means: Extending k-means with efficient estimation of the number of clusters,” in Proceedings of the Seventeenth International Conference on Machine Learning, , 2000 pp. 727–734.

[9] G. Celeux and G. Govaert, “A classification EM algorithm for clustering and two stochastic versions,” Computational statistics and data analysis, vo.14, pp.315–332, 1992.

[10] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, “CDnet 2014: An Expanded Change Detection Benchmark Dataset,” in Proceedings of the IEEE Workshop on Change Detection (CDW-2014), 2014.